Molecular evolution of growth hormone

Article (Accepted Version)

Article for *The Biochemist*. Special Issue on *Molecular Evolution*

MOLECULAR EVOLUTION OF GROWTH HORMONE

Michael Wallis

Biochemistry and Molecular Biology Group, School of Life Sciences, University of Sussex, UK)

tel: 01273 472552; email: m.wallis@sussex.co.uk

Growth hormone (GH) is a single-chain protein hormone of about 190 residues, expressed mainly in the vertebrate anterior pituitary gland, which regulates somatic growth and various aspects of metabolism; many of these actions are mediated by insulin-like growth factor 1 (IGF1). Growth defects in humans frequently result from GH deficiency and are often treatable by GH administration. The evolution of GH illustrates many features of molecular evolution, including (i) the development and elaboration of gene/protein families by gene duplication, (ii) subtle changes resulting from incorporation of point mutations, which often occur during episodes of accelerated change, and (iii) co-evolution of hormones and their receptors.

GH is structurally related to another pituitary hormone, prolactin, which in mammals regulates mammary growth and lactation and in lower vertebrates exhibits a wide range of activities, many relating to secondary aspects of reproduction. Neither GH nor prolactin has been convincingly shown to occur in invertebrates. The biological actions of GH and prolactin show considerable species specificity, as do their primary structures and gene organization, though linking biological and molecular variation is often difficult[1].

Gene duplication in the evolution of GH and prolactin

GH and prolactin are structurally similar proteins at both the sequence level (about 25% sequence identity) and at the level of 3-dimensional structure (both are 4-helix bundles with an unusual "up-up-down-down" topology - Figure 1). They are also encoded by similar genes, each with 5 exons and 4 introns, although the prolactin gene is larger (~10kb) than that of GH (~2 kb). The two genes are found on different

chromosomes in man, but clearly arose as a result of gene duplication followed by divergent evolution. The duplication probably occurred early in vertebrate evolution - GH and prolactin are distinct hormones in all but the most primitive vertebrates, the cyclostomes. A related protein, somatolactin, is found in many fish, with about equal identity to GH and prolactin, presumably reflecting an additional gene duplication.

A number of proteins, including erythropoietin and many interleukins, are distantly related to GH and prolactin and comprise a cytokine protein superfamily. Their structures all show the unusual up-up-down-down 4-helix bundle found in GH and prolactin (Figure 1), and their receptors also show similarity, with a single transmembrane domain and some conserved motifs. This cytokine family probably arose as a consequence of gene duplications that occurred much earlier than that giving rise to GH and prolactin, before or soon after the origin of vertebrates. The period of divergent evolution that followed led to the loss of most sequence similarity between the cytokines, though the characteristic 3D fold was retained (3D structure is usually more conserved than sequence).

More-recent duplications of prolactin and GH genes occurred during mammalian and avian evolution. In passerine birds, duplication of the GH gene, was followed by rapid evolution[2]. Both genes are expressed in zebra finch brain, but their functions are not clear. In mammals the gene duplications have given rise to families of proteins expressed mainly in the placenta. In ruminant artiodactyls and in rodents, independently, repeated duplications of the prolactin gene gave rise to up to 20 closely-linked genes. The protein products include placental lactogens (PLs) but their functions remain incompletely understood.

Higher primates also have a comparable cluster of genes expressed in the placenta, but these derive from multiple duplications of the GH gene. In human five GH-like genes are clustered on chromosome 17[3] (Figure 2). The *GH-N* gene encodes pituitary GH, while the other genes are expressed in the placenta. Two genes (*PL-A* and *PL-B*) encode identical sequences for PL, about 85% identical to GH-N, expressed at high levels during pregnancy. *PL-L* (placental lactogen-like) encodes an abnormally spliced mRNA and may be a functionless pseudogene. *GH-V* encodes a GH variant, about 93% identical to GH-N, which is expressed in the placenta at moderate levels, apparently taking over from pituitary GH during pregnancy. Human GH is much more similar to the human PL than to non-primate GHs, and phylogenetic analysis confirms that the gene duplications that gave rise to this gene cluster occurred during the course of primate evolution. The role of human PL is not fully established. Experimentally they have lactogenic activity and relatively low growth-promoting activity, but they disappear from the circulation soon after parturition, at the time when lactogenesis increases. In some rare cases parturition and subsequent lactation are normal despite lack of *PL* genes and circulating PL. PLs may play a role in regulating the metabolic balance between mother and foetus when nutritional resources are limited. Maintaining this balance could involve maternal-foetal competition - the *PL* and *GH-V* genes are expressed on the foetal side of the placenta.

Investigation of GH-related genes in other primates has thrown light on the evolution of this gene cluster. In lower primates (including slow loris and bushbaby) there is a single *GH* gene, as in most non-primate mammals. Monkeys and apes have a cluster of *GH*-related genes, as in man (Figure 2)[4-6]. Remarkably, the GH-gene clusters found in Old

World Monkeys (OWM)/apes and New World Monkeys (NWM) (Figure 2) have arisen independently[1,4,7]. Each of these clusters contains several genes, at least some of which are expressed in the placenta, but their separate origins are shown by phylogenetic analysis, distribution of repetitive elements (*Alu*), and identification of "break-points" for the initial duplications. In one NWM, the capuchin, *Cebus albifrons*, the cluster comprises at least 40 *GH*-like genes and (mostly) pseudogenes[4] - gene duplication seems to have run amok. The evolutionary forces that led to independent duplications of the *GH* gene in both OWM/apes and NWM are not clear, but it is notable that they followed a period of rapid evolution of GH (see next section).

Episodic evolution of GH

A conventional view of molecular evolution holds that for any given protein the rate of evolution is rather constant, but that rates for different proteins vary considerably. In contrast, GH evolution shows a pattern in which slow, apparently rather constant, evolutionary change (near stasis) is interrupted by short episodes during which the rate of evolution increased 10-100-fold. This episodic evolution is seen for amino acid changes/nonsynonymous nucleotide substitutions but not for synonymous substitutions which do not affect protein structure (Figure 3). This and other evidence suggests that, in some cases at least, the accelerated evolution was adaptive and associated with subtle changes in biological activity.

A marked episode of accelerated GH evolution occurred on the lineage leading to higher primates, giving a substantial sequence difference (at about 35% of all residues) between human GH and GHs of non-primate mammals. This explains why

non-primate GHs are not active in man so that, before recombinant DNA-derived human GH became available, human GH deficiency had to be treated with GH extracted from human pituitary glands collected *post-mortem*. The specific changes in both GH and its receptor that underlie the specificity arising during this burst of rapid change have been identified[8], though changes in species specificity cannot themselves underlie the adaptive changes driving the accelerated evolution. A possible mechanism explaining the episode of rapid evolution is "function switching", in which a protein has two or more biological actions, the relative importance of which varies over the course of evolutionary time[1]. Repeated alternating adaptation of the hormone to performance of each of the functions would lead to accumulation of many amino acid substitutions with rather little overall change in biological activity. The pressure for rapid change could stop following the gene duplication, with one of the duplicate genes retaining one function of the hormone and the second adopting the other(s). A relevant second function for GH could have involved placental expression, which may have had an early origin in primate evolution[6]. Notably, the episode of rapid evolution *preceded* the gene duplications that gave rise to GH-gene clusters in OWM/apes and NWM (see above), and after gene duplication the rate of evolution of GH decreased. This, and the observation that the burst of rapid change is seen for nonsynonymous substitutions/amino acid changes in mature GH, but is much less marked for synonymous substitutions (Figure 3), or changes in signal peptides, 5' sequences or introns, supports the view that the changes during the accelerated evolution were adaptive in nature.

Further episodes of rapid GH evolution are seen on several other mammalian lineages[9], including the branch leading to ruminant artiodactyls (Figure 3). Among

lower vertebrates the rate of evolution is again generally slow (remarkably, the sequence of GH from a non-primate mammal such as pig is more similar to that of a shark than to human GH). However, rates of evolution of teleost GHs are variable, and in some cases very high.

Coevolution of GH, prolactin and their receptors

Like all polypeptide hormones, GH and prolactin act by binding to receptors on the plasma membranes of their target cells. The similarity between the two hormones is reflected in a corresponding similarity between the receptors. Presumably the gene duplication followed by divergence that gave rise to GH and prolactin was paralleled by a corresponding duplication and divergence giving rise to their receptors. As a consequence there is some overlap in the receptor-binding and biological actions of the hormones - for example, in some species, including human, GH has quite high lactogenic activity. Such overlap suggests the possibility of coevolution of GH and prolactin, and the episodic evolution of the hormones at least partly accords with this. Like GH, prolactin evolution in mammals shows an episodic pattern and in some cases bursts of rapid evolution of prolactin coincide with those seen for GH (e.g. on lines leading to higher primates and ruminants), though in other cases there is no such coincidence[10].

Coevolution is also seen between GH and prolactin and their corresponding receptors. Thus, the episode of rapid change of GH during primate evolution is accompanied by a corresponding burst of evolution of the GH receptor. The changes in the receptor are largely confined to the extracellular domain (the region that interacts with the

hormone) and, for both hormone and receptor, residues close to the hormone-receptor binding site are particularly involved (Figure 4). On the other hand, for the accelerated evolution of GH seen on the lineage leading to armadillo (Figure 3), most changes occurred on the side of the hormone distant from the receptor-binding site (Figure 4), suggesting interaction of the hormone-receptor complex with one or more additional proteins[9].

Conclusion: tempo and mode in the evolution of protein hormones

The evolution of GH displays an unusual pattern of molecular evolution. For much of evolutionary time the hormone has been strongly conserved - a state of near stasis. But for brief periods rapid molecular change occurred, with change in the sequence of the hormone and/or the number of genes. During most of GH evolution in mammals (80-90%) the condition of near-stasis applied, but most of the evolutionary change (60-70%) occurred during short bursts corresponding to 10-20% of evolutionary time. In some cases the episodes of rapid change coincide with those seen for prolactin and/or the GH receptor, suggesting coevolution.

Such a pattern disagrees with general expectation, and attempts have been made to explain it in part as reflecting *GH* gene gain and loss on various lineages[6]. Such an explanation seems unnecessarily complicated, particularly since detailed scrutiny of mammalian genomes provides no indication of such gain and loss, and the discovery of additional lineages showing apparent accelerated evolution of GH within mammals[9] would imply additional undetected gain and loss. In those cases where

duplications of the *GH* or prolactin gene have occurred, they have followed accelerated evolution rather than preceded it.

Is this pattern of bursts of rapid change occurring within an overall near-stasis confined to GH, or might it apply more generally? Scrutiny of other protein hormones from a fairly limited number of mammalian species suggested that about half of them showed an episodic pattern of this sort[10], including insulin, the common alpha-subunit of the glycoprotein hormones, and the beta-subunit of luteinizing hormone. A similar pattern is seen in several other proteins, including haemoglobin and cytochrome c. Although it was at one time thought that most evolutionary change in proteins was neutral in nature, it now seems likely that a substantial proportion of such change - perhaps 50% - is adaptive. It may be that a large proportion of this adaptive change occurs in episodes of accelerated evolution.

REFERENCES

1. Forsyth, I.A. and Wallis, M. (2002). J. Mammary Gland Biol. Neoplasia **7**, 291-312.
2. Yuri, T., Kimball, R.T., Braun, B.L. and Braun, M.J. (2008) Mol. Biol. Evol. 25, 352-361.
3. Chen EY, Liao YC, Smith DH et al. (1989) Genomics **4**, 479–497.
4. Wallis, O.C. and Wallis, M. (2006). J. Mol. Evolution **63**, 591–601.
5. Pérez-Maya, A.A., Rodríguez-Sánchez, I.P., de Jong, P., Wallis, M. and Barrera-Saldaña, H.A. (2012) Mammalian Genome **23**, 387-398.
6. Papper, Z., Jameson, N.M., Romero, R. et al. (2009). Proc. Nat. Acad. Sci. U.S.A.

106, 17083–17088.

7. Li, Y., Ye, C., Shi, P. et al. (2005). J. Mol. Endocrinol. **21**, 1-5.

8. Liu, J.C., Makova, K.D., Adkins, R.M., Gibson, S. and Li, W.-H. (2001) Mol. Biol. Evol. **18**, 945–953.

9. Wallis, M. (2008) Gen. Comp. Endocrinol. **155**, 271-279.

10. Wallis, M (2001) J. Mol. Evol. **53**, 10-18.

FIGURE LEGENDS

Figure 1. Structures of human GH (a) and Prolactin (b). Like other members of the cytokine superfamily the structure of these hormones is dominated by a 4-helix bundle; the conformation of this (up-up-down-down) differs from that seen in most 4-helix bundle proteins (up-down-up-down) as illustrated in (c). Structural representations were constructed using PyMol and pdb entries 3hhr chain A (GH) and 1RW5 (prolactin).

Figure 2. The organization of *GH* gene clusters in primates. In prosimians such as the slow loris, as in most non-primate mammals, there is a single *GH*-like gene, expressed in the pituitary gland. In higher primates there is a complex cluster of *GH*-like genes, with pituitary-expressed *GH* at the 5' end followed by several placentally expressed genes (including those for placental lactogens, PLs). The clusters arose by sequential tandem duplications, apparently independently in New World monkeys and Old World monkeys/apes. Many of the clusters include one or more non-functional pseudogenes.

Figure 3. Evolutionary trees for mammalian GHs. Trees were derived from coding sequences for mature GHs from representative species, using the codeml programme. Separate trees are shown, based on non-synonymous substitutions (A, essentially equivalent to amino acid sequences) and synonymous substitutions (B). In A, evolutionary rate is markedly variable, with episodes of rapid change occurring on the branches marked with thick lines; this contrasts with the much less variable evolutionary rates seen in B. Evolutionary time is represented on the abscissa.

Numbers on branches are numbers of substitutions; the approximate times of duplications that gave rise to the GH-related placental proteins in NWM and OWM are shown by ◆.

Figure 4. Binding of GH to the extracellular domain of its receptor. GH is shown in space-filling format (blue and yellow), while the two chains of the receptor are shown in line format (purple). For each model three views are given, (a) sideways on (membrane at the bottom), (b) from the top and (c) from the bottom (looking up from the membrane). Residues which changed during the episodes of accelerated change occurring in the evolution of armadillo GH and human GH (Figure 3) are shown in yellow. For armadillo GH (top) most of the substitutions occurred in the region of the molecule away from the receptor. In human GH (bottom) most substitutions occurred on the side close to the membrane, with a substantial proportion (~30%) within 5A of the receptor-binding site. The difference is statistically significant. Constructed using PyMOL and pdb entry 3hhr (human GH bound to the extracellular domain of its receptor).

Fig. 1

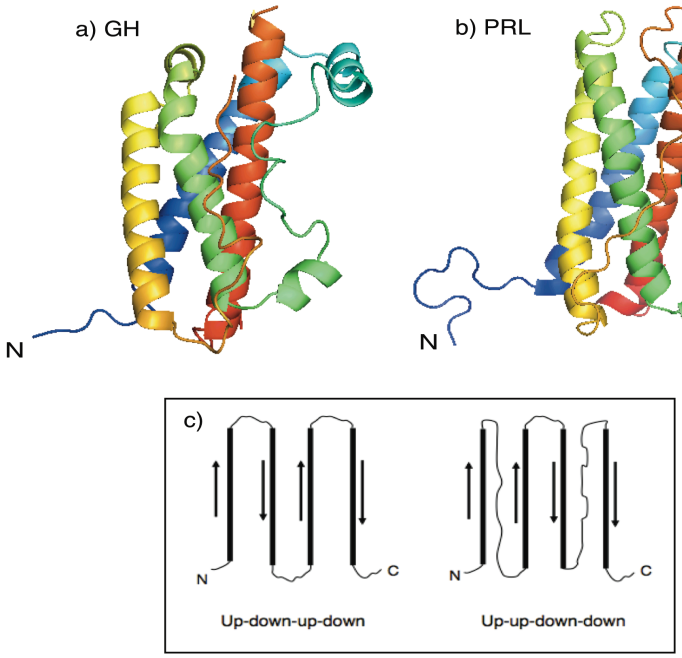

a) GH
b) PRL
N
N
c)
N
C
N
C
Up-down-up-down
Up-up-down-down

Fig. 2

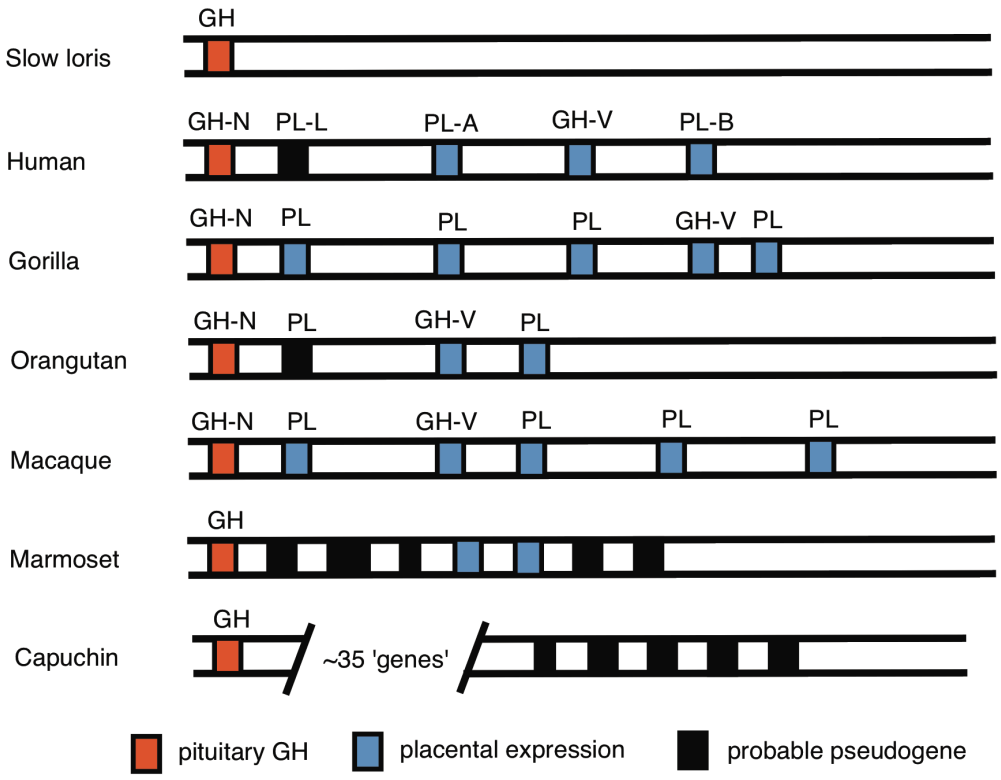

Slow loris
GH
Human
GH-N
PL-L
PL-A
GH-V
PL-B
Gorilla
GH-N
PL
PL
PL
GH-V
PL
Orangutan
GH-N
PL
GH-V
PL
Macaque
GH-N
PL
GH-V
PL
PL
PL
Marmoset
GH
Capuchin
GH
~35 'genes'
pituitary GH
placental expression
probable pseudogene

Fig. 3

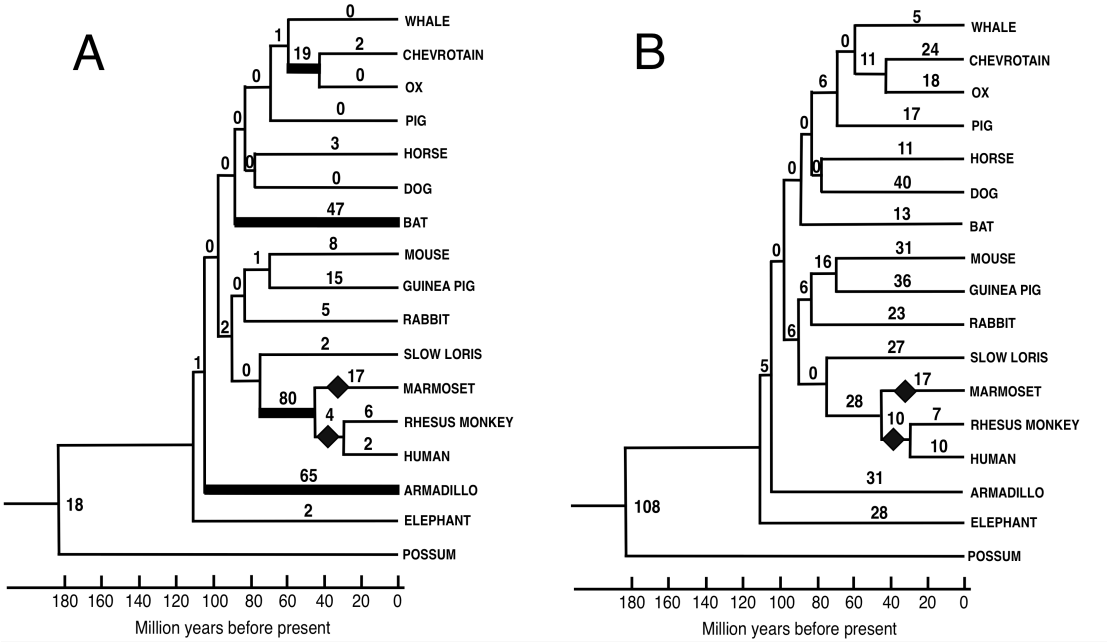
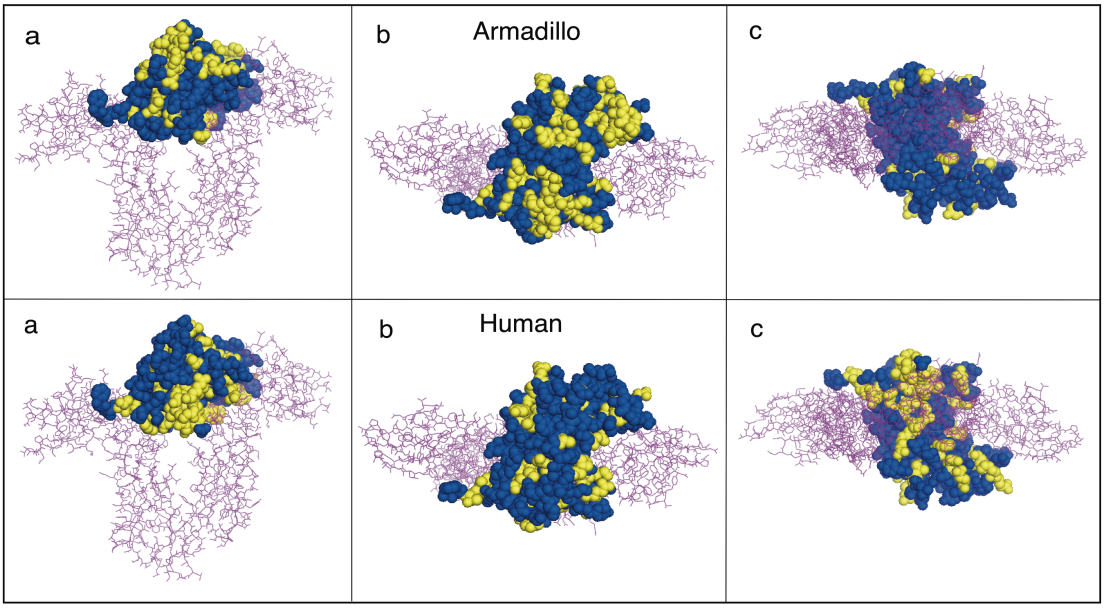

A
WHALE
CHEVROTAIN
OX
PIG
HORSE
DOG
BAT
MOUSE
GUINEA PIG
RABBIT
SLOW LORIS
MARMOSET
RHESUS MONKEY
HUMAN
ARMADILLO
ELEPHANT
POSSUM
180 160 140 120 100 80 60 40 20 0
Million years before present
B
WHALE
CHEVROTAIN
OX
PIG
HORSE
DOG
BAT
MOUSE
GUINEA PIG
RABBIT
SLOW LORIS
MARMOSET
RHESUS MONKEY
HUMAN
ARMADILLO
ELEPHANT
POSSUM
180 160 140 120 100 80 60 40 20 0
Million years before present

Fig. 4


a
b
Armadillo
c
a
b
Human
c